B. Biterge Süt / Hacettepe J. Biol. & Chem., 2021, 49 (4), 367-374

Research Article

Hacettepe Journal of Biology and Chemistry

journal homepage: www.hjbc.hacettepe.edu.tr

Structural Analysis of Novel Amino Acid Substitutions in SARS-CoV-2 Spike Protein Receptor-Binding Domain

SARS-CoV-2 Spike Protein Reseptör-Bağlanma Bölgesindeki Aminoasit Değişimlerinin Yapısal Analizi

Burcu Biterge Süt

Department of Medical Biology, Faculty of Medicine, Niğde Ömer Halisdemir University, Niğde, Turkey.

ABSTRACT

There are several novel amino acid substitutions in SARS-CoV-2 spike protein, which could account for the increased infectivity of this newly emerged virus. Therefore, in this paper we aimed to evaluate the potential effects of these amino acid substitutions on protein structure and function. For this purpose, we made use of several state-of-the-art computational tools and performed in silico analyses on protein similarity, 2D and 3D structure, ligand binding and biological function. We found that some of the novel amino acid changes caused significant structural alterations both at the secondary and tertiary structure level, possibly affecting the interaction between the spike protein receptor-binding domain (RBD) and ACE2, as well as other ligands. In conclusion, data we provided here is a significant contribution to our current knowledge of the SARS-CoV-2 virus and will aid in having a better understanding of its molecular differences, mechanism of infection and the cellular processes it affects in the host in order to develop better therapies and vaccines.

Key Words

COVID-19, coronavirus, SARS-CoV-2, spike protein, RBD, structural analysis.

ÖZ

SARS-CoV-2 spike proteininde, bu yeni ortaya çıkan virüsün enfeksiyöz özelliğindeki artışı açıklayabilecek birçok aminoasit değişimi mevcuttur. Bu nedenle, bu çalışmada, söz konusu aminoasit değişimlerinin protein yapısı ve fonksiyonu üzerindeki potansiyel etkisinin değerlendirilmesi hedeflenmiştir. Bu amaçla, gelişmiş teknoloji ürünü bilişimsel araçlar kullanılarak protein benzerliği, 2D ve 3D yapı, ligand bağlanması ve biyolojik fonksiyon üzerine in siliko analizler yapılmıştır. Bu aminoasit değişimlerinden bazılarının iki- ve üç-boyutlu protein yapısında önemli değişimlere neden olduğu ve spike proteini reseptör-bağlanma bölgesi (RBB) ile ACE2 ve diğer ligandlar arasındaki etkileşimi değiştirebileceği belirlenmiştir. Sonuç olarak, bu çalışmada sunulan veriler SARS-CoV-2 virüsüne yönelik halihazırda sahip olduğumuz bilgi dağarcığına önemli bir katkı sağlayarak; virüsün moleküler farklılıklarının, enfeksiyon mekanizmasının ve konak hücrede etkilediği hücresel süreçlerin daha iyi anlaşılması ile daha etkili tedavi ve aşıların geliştirilebilmesine yardımcı olacaktır.

Anahtar Kelimeler

COVID-19, koronavirüs, SARS-CoV-2, spike protein, RBB, yapısal analiz.

Article History: Received: Aug 20, 2020; Revised: Feb 9, 2021; Accepted: Feb 25, 2021; Available Online: Jun 10, 2021.

DOI: https://doi.org/10.15671/hjbc.776430

Correspondence to: B. Biterge Süt, Department of Medical Biology, Faculty of Medicine, Niğde Ömer Halisdemir University, Niğde, Turkey.

E-Mail: bbitergesut@ohu.edu.tr

INTRODUCTION

The recent outbreak of Coronavirus Disease-2019 (COVID-19) is caused by a novel and highly pathogenic coronavirus (severe acute respiratory syndrome coronavirus-2, SARS-CoV-2) [1] and quickly evolved into a global health concern. Infections of Coronaviridae, which constitute a family of vertebrate-infecting RNA viruses [2], have previously been established to result in common colds in humans and were associated with the spread of SARS in 2003 [3]. There are at least seven different coronaviruses identified that cause human diseases including MERS (Middle East respiratory syndrome, MERS-CoV), SARS (SARS-CoV) and COVID-19 [4].

SARS-CoV-2 has an unusually large genome, which is almost 30 kb. It has six major open reading frames (ORFs) and encodes for four structural proteins S (spike), E (envelope), M (membrane) and N (nucleoprotein), as well as sixteen non-structural proteins (Nsp1-16) [5]. Spike protein is critically important for viral infectivity, as it binds to the human cell surface receptor angiotensin-converting enzyme 2 (ACE2) and mediates the entry of the virus into the host cell [6]. Structurally, spike protein contains a cleavage site giving rise to two subunits namely S1 and S2 (Figure 1.a), which play a sequential role in mediating the viral entry and the completion of membrane fusion with the host [7, 8]. The receptor-binding motif (RBM), which lies within the receptor-binding domain (RBD) of the S1 subunit, provides contact with the ACE2 receptor. S2 subunit includes two heptad repeats HR1 and HR2, ensuring proper translocation of the viral genome into the infected cell [9]. Due to its key role in receptor recognition, Spike protein is widely considered as a promising target for raising specific antibodies against and developing vaccines [10].

Phylogenetically, viruses of the MERS, SARS and COVID-19 diseases belong to the same genus and utilize similar entry pathways. While MERS-CoV is more distantly related to SARS-CoV and SARS-CoV-2, the latter two share significant similarity. Despite this substantial conservation, several mutations in SARS-CoV-2 spike and nucleocapsid proteins have been revealed by structural analyses [11]. A better understanding of these genetic variations is crucial, as they could account for the enhanced infectivity of the COVID-19 virus in comparison to the SARS coronavirus. Therefore, this study aims to evaluate the potential effects of novel amino acid substitutions in SARS-CoV-2 spike protein receptor-binding domain on protein structure and function.

MATERIALS and METHODS

Determination of sequence similarity

The amino acid sequences of spike protein for SARS-CoV (ID: P59594) and SARS-CoV-2 (ID: P0DTC2) were retrieved from UniProt. Sequence similarity was evaluated using Clustal Omega (https://www.ebi.ac.uk/Tools/msa/clustalo/) [12]. Clustal Omega calculates a similarity score based on Gonnet PAM 250 matrix and predicts conservation as strong for values above 0.5; denoted by a colon (:), while values below 0.5 are considered weakly similar; indicated by a period (.). An asterisk (*) refers to a fully conserved residue.

Structural analyses

In silico prediction of 2D and 3D structures of the SARS-CoV and SARS-CoV-2 spike proteins was performed via I-TASSER (https://zhanglab.ccmb.med.umich.edu/) [13]. I-TASSER deduces the helix-strand-coil structure of a queried protein, builds 3D models and predicts potential ligands and ligand binding sites based on these models provided with a confidence score (C-score). C-scores vary between (-5) to 2, a higher value indicating higher confidence. For further analyses, models with the highest C-scores were selected.

3D models obtained from I-TASSER were visualized and superimposed using UCSF Chimera (https://www.cgl.ucsf.edu/chimera/) [14]. Side chains were denoted for the residues that displayed structural differences.

Functional analyses

Gene Ontology (GO) terms for molecular function, biological process and cellular component were predicted by and retrieved from I-TASSER. List of consensus GO terms was created by selecting the top GO terms with the highest confidence values. Details regarding the GO terms were acquired from Quick GO (https://www.ebi.ac.uk/QuickGO/).

Statistical analysis

All confidence scores were calculated and statistical analyses were performed automatically by the respective tool used for the analysis.

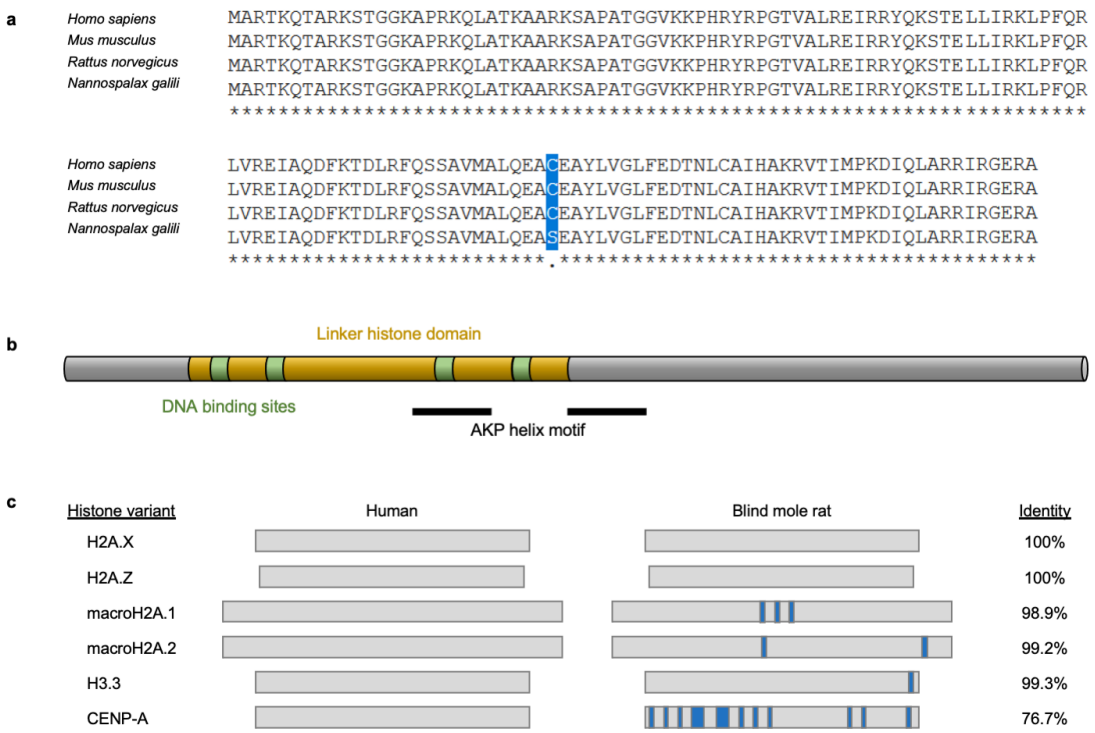

Figure 1. Domain structure of the SARS-CoV-2 spike protein (a). Receptor binding domain (RBD, shown in yellow and green) is significantly similar between the two viruses (b). Novel amino acid substitutions in SARS-CoV-2 spike protein receptor-binding domain results in alterations of the 2D structure between residues ~110-194 (c).

RESULTS

Conservation of the receptor-binding domain (RBD)

Receptor-binding domain of the SARS-CoV-2 spike protein is a 194 amino acid-long stretch located within the S1 subunit (Figure 1.a). Both the full-length spike protein and the receptor-binding domain presented significant similarity between SARS-CoV and SARS-CoV-2. However, for simplicity, we focused on the receptor-binding domain for further analyses and showed that 143 amino acids (73.7 %) were an identical match (Figure 1.b). 22 amino acids (11.3 %) were strongly and 12 amino acids (6.2 %) were weakly similar; whilst, 17 amino acids (8.8%) did not share any similar properties.

Effects on structural configuration

The secondary (2D) structure of a protein is an intermediary step before completion of protein-folding into the tertiary (3D) structure. The 2D structure is composed of α-helices, β-sheets (strands) and random coils that are held together by hydrogen bonds. The prediction of the RBD secondary structure indicated the formation of helices, strands and coils at the same residues for both viruses for approximately the first 110 amino acids (Figure 1.c). However, it was immediately followed by an additional β-strand in SARS-CoV-2 RBD. The next three β-strands seem to have shifted either backward or forward, while part of the strand at the very C-terminus was converted to a helix.

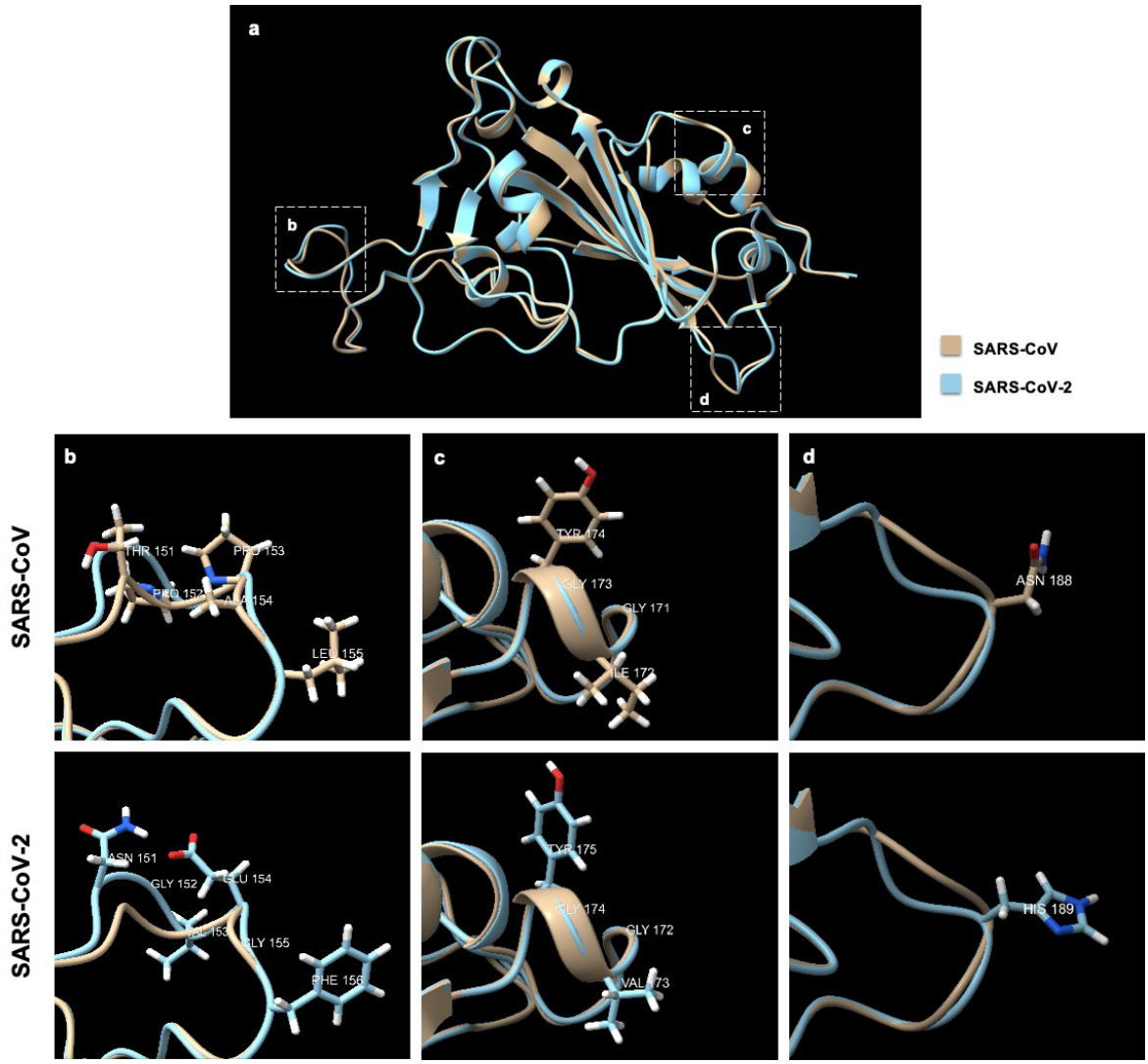

Figure 2. 3D constructions and superimposition of the SARS-CoV and SARS-CoV-2 spike protein RBDs exhibit conserved structures (a). Residues that displayed different structural configurations are marked by dashed squares and zoomed-in views are shown in the lower panels (b, c, d). Side chains are denoted for the marked residues.

The tertiary structures of the RBDs were mostly identical and superimposed with a high level of similarity (Figure 2.a). The novel amino acid substitutions in SARS-CoV-2 spike protein RBD did not result in major structural alterations except for three regions between amino acids 151-155/6, 171/2-174/5 and 188/189 (note that there is a shift in amino acid positions due to a missing SARS-CoV amino acid at position 153). The TPPAL residue (aa151-155) in SARS-CoV was drastically mutated into NGVEGF (aa151-156) in SARS-CoV-2, altering the 3D conformation (Figure 2.b). In Figure 2.c, a single missense mutation of I172 to V173 caused the strand formation of a four amino acid-long residue to change into a loop. Similarly, the conversion of N188 to H189 considerably affected the tertiary structure of a larger residue (Figure 2.d).

Functional implications

Structural variations within a protein or a domain often pose functional significances as well. Regardless of how subtle the variation seems, it could change interactions between the protein and its substrate, ligand or receptor. In line with this, we identified different sets of potential ligands for SARS-CoV and SARS-CoV-2 RBDs (Table 1). ATP, biotinol-5-AMP, alpha-D-mannose, riboflavin monophosphate and uridine-5'-diphosphate were predicted as the top potential ligands for SARS-CoV, whereas ligands for SARS-CoV-2 were predicted

Table 1. Ligands and ligand binding sites of SARS-CoV and SARS-CoV-2 receptor binding domains.

	Rank	C-score	Ligand Name	Ligand Binding Site Residues (aa)
SARS-CoV RBD	1	0.07	ATP	21, 22, 23, 24, 69, 71, 121
	2	0.07	Biotinol-5-AMP	8, 34, 101, 102, 103, 181, 182, 183, 185
	3	0.05	Alpha-D-Mannose	26, 178
	4	0.05	Riboflavin Monophosphate	71, 109, 176
	5	0.02	Uridine-5'-Diphosphate	12, 14, 17, 18
SARS-CoV-2 RBD	1	1	Alpha-D-Mannose	26, 179
	2	2	Oxalate $^{(2-)}$	29
	3	3	ADP	22, 23, 24, 69, 70, 71, 122
	4	4	K-252A	6, 8, 36, 58, 59, 60, 61, 64, 185
	5	5	N-acetylglucosamine	24, 71, 179

as alpha-D-mannose, oxalate$^{(2-)}$, ADP, K-252A (a serine/threonine kinase inhibitor) and N-acetylglucosamine. Figure.3 shows the 3D conformation of SARS-CoV and SARS-CoV-2 RBDs in complex with their predicted ligands with the highest confidence values, ATP and alpha-D-mannose, respectively.

Gene Ontology (GO) enrichment analysis of SARS-CoV and SARS-CoV-2 RBDs identified different but somewhat overlapping molecular functions and biological processes (Table 2). Among these, ATP and metal/ion binding, ligase and transferase activities, as well as tRNA-related mechanisms are prominent. Furthermore, SARS-CoV and SARS-CoV-2 RBDs were predicted to differ in their localization to cellular components; SARS-CoV RBD localizes to the cytoplasm while SARS-CoV-2 RBDs localizes to the intracellular part, which includes both the cytoplasm and the nucleus.

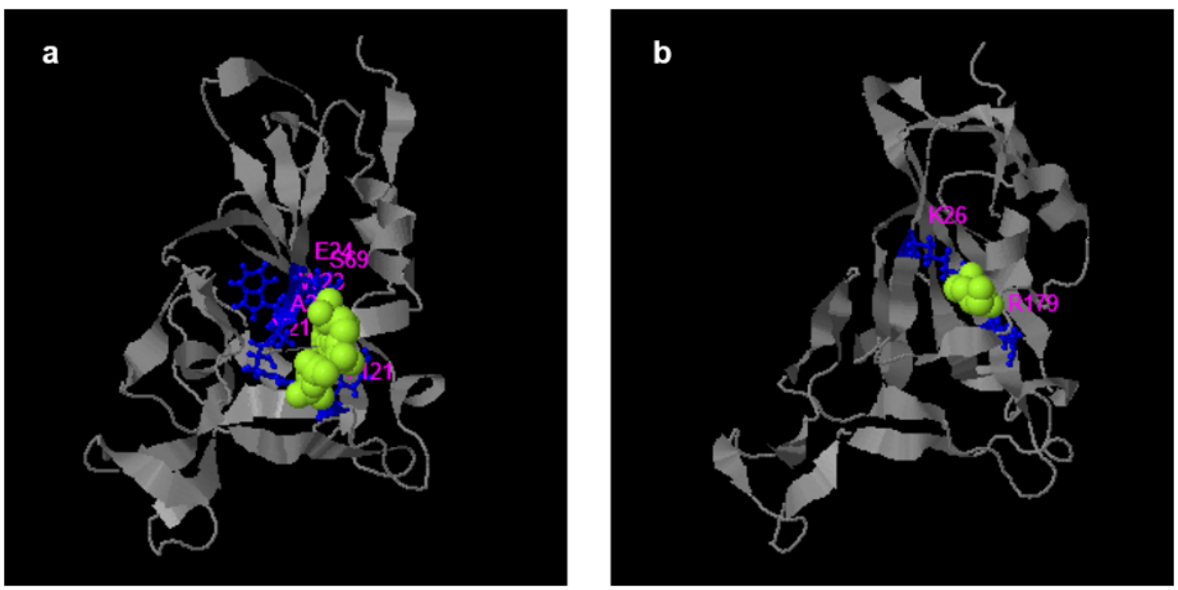

Figure 3. 3D conformation of SARS-CoV (a) and SARS-CoV-2 (b) RBDs in complex with their predicted ligands ATP and alpha-D-mannose (in green), respectively. Residues involved in ligand binding are indicated.

Table 2. Consensus prediction of SARS-CoV and SARS-CoV-2 RBD related GO terms.

	SARS-CoV RBD		SARS-CoV-2 RBD	
	GO term ID	GO term	GO term ID	GO term
Molecular Function	GO:0046872	Metal ion binding		
	GO:0016879	Ligase activity, forming carbon-nitrogen bonds	GO:0005524	ATP binding
	GO:0016830	Carbon-carbon lyase activity	GO:0043167	Ion binding
	GO:0016740	ATP binding	GO:0016884	Carbon-nitrogen ligase activity
	GO:0005524	Transferase activity	GO:0016740	Transferase activity
	GO:0016876	Aminoacyl-tRNA ligase activity		
Biological Process	GO:0009292	Genetic transfer		
	GO:0000746	Conjugation		
	GO:0044272	Sulfur compound biosynthetic process		
	GO:0009108	Coenzyme biosynthetic process	GO:0009292	Genetic transfer
	GO:0009106	Lipoate metabolic process	GO:0000746	Conjugation
	GO:0018130	Heterocycle biosynthetic process	GO:0018065	Protein-cofactor linkage
	GO:0046394	Carboxylic acid biosynthetic process	GO:0006412	Translation
	GO:0018065	Protein-cofactor linkage		
	GO:0006399	tRNA metabolic process		
	GO:0034645	Cellular macromolecule biosynthetic process		
Cellular Component	GO:0005737	Cytoplasm	GO:0044424	Intracellular part

DISCUSSION

It has been more than six months since the first appearance of the novel coronavirus SARS-CoV-2; however, effective antiviral therapies and vaccines against it are yet to be developed and approved. Currently, the disease is mostly managed by supportive care, which is greatly limited by the boundaries of the healthcare systems available. Therefore, it is immensely important to focus on developing targeted therapies to contain the further spread of the virus. Spike protein is an obvious candidate in this respect, as it is critical for infecting the host cell. Studies have suggested that specific antibodies against receptor-binding domain of the spike protein could effectively block the viral entry [15]. In this study, we identified novel amino acid substitutions in SARS-CoV-2 spike protein receptor-binding domain and evaluated their potential effects on protein structure and function.

We showed that although the SARS-CoV and SARS-CoV-2 spike proteins are highly similar, there were novel amino acid substitutions within the RBD. Some of these amino acid changes caused significant structural alterations both at the secondary and tertiary structure level although the others did not. We observed that all amino acid substitutions introduced distinct side chains, which are important determinants of hydrogen bond formation and establishment of the 2D and 3D structures. For instance, it was shown that due to the substitution of P145 in SARS-CoV RBD to A145 in SARS-CoV-2 RBD, an additional hydrogen bond forms between A145 and N157; compacting the protein and moving the binding site closer to ACE2 [16]. Furthermore, the V87K substitution was shown to cause the formation of a salt bridge between SARS-CoV-2 RBD and ACE2, which cannot form in SARS-CoV RBD [8]. Additionally, we identified three regions between amino acids 151-155/6, 171/2-174/5 and 188/189, which resulted in major structural alterations. Among these residues, E154, G155, F156 and V173 are unique to SARS-CoV-2 RBD and provide additional sites for interacting with ACE2 [17]. This increase in the number of ACE2-interacting amino acids could suggest increased binding affinity of SARS-CoV-2 RBD in comparison to SARS-CoV RBD, explaining its enhanced infectivity.

Studies have proposed that the interaction with ACE2 can also be affected by other factors such as ligands, in addition to the structural alterations in the SARS-CoV-2 RBD [18]. This fits well with our observation that SARS-CoV and SARS-CoV-2 RBDs interact with different sets of potential ligands. Moreover, we showed that both RBDs are implicated in several important molecular functions and biological processes. Although the actual involvement of the spike protein RBDs in these cellular processes needs experimental evidence, in support of this, a previous study showed that the coronavirus spike protein induces endoplasmic reticulum stress [19].

In conclusion, the fight against the Coronavirus Disease-2019 is not over yet. In order to develop better therapies and vaccines, it is vital to have a better understanding of the molecular differences in this novel type of coronavirus, its mechanism of infection and the cellular processes it affects in the host. Our ***in silico*** evaluation of the novel amino acid substitutions in SARS-CoV-2 spike protein receptor-binding domain from a structural and functional point is a significant contribution to the literature.

Acknowledgments – None.

References

1. Q. Li, X. Guan, P. Wu, X. Wang, L. Zhou, Y. Tong et al, Early transmission dynamics in Wuhan, China, of novel coronavirus-infected pneumonia, N. Engl. J Med., 382(2020) 1199-1207.
2. P.S. Masters, The molecular biology of coronaviruses, Adv. Virus Res., 66 (2006) 193-292.
3. C. Drosten, S. Günther, W. Preiser, S. van der Werf, H.R. Brodt, S. Becker et al, Identification of a novel coronavirus associated with severe acute respiratory syndrome, N. Engl. J. Med., 348 (2003) 1967-1976.
4. J. Zheng, SARS-CoV-2: an emerging coronavirus that causes a global threat, Int. J. Biol. Sci., 16 (2020) 1678-1685.
5. D.E. Gordon, G.M. Jang, M. Bouhaddou, J. Xu, K. Obernier, K.M. White et al, A SARS-CoV-2 protein interaction map reveals targets for drug repurposing [published online ahead of print]. Nature. (2020) doi:10.1038/s41586-020-2286-9.
6. H. Zhang, J.M. Penninger, Y. Li, N. Zhong, A.S. Slutsky. Angiotensin-converting enzyme 2 (ACE2) as a SARS-CoV-2 receptor: molecular mechanisms and potential therapeutic target, Intensive Care Med., 46 (2020) 586-590.
7. G. Simmons, P. Zmora, S. Gierer, A. Heurich, S. Pöhlmann, Proteolytic activation of the SARS-coronavirus spike protein: cutting enzymes at the cutting edge of antiviral research, Antiviral Res., 100 (2013) 605-614.
8. J. Lan, J. Ge, J. Yu, S. Shan, H. Zhou, S. Fan, et al, Crystal structure of the 2019-nCoVspike receptor-binding domain bound with the ACE2 receptor, J. Bio. Rxiv., (2020) http://doi.org/10.1101/2020.02.19.956235%JbioRxiv.
9. W.E. Chan, C.K. Chuang, S.H. Yeh, M.S. Chang, S.S. Chen. Functional characterization of heptad repeat 1 and 2 mutants of the spike protein of severe acute respiratory syndrome coronavirus, J. Virol., 80 (2006) 3225-3237.

10. G. Salvatori, L. Luberto, M. Maffei, L. Aurisicchio, G. Roscilli, F. Palombo et al, SARS-CoV-2 spike protein: an optimal immunological target for vaccines, J. Transl. Med., 18 (2020) 222.
11. D. Benvenuto, M. Giovanetti, A. Ciccozzi, S. Spoto, S. Angeletti, M. Ciccozzi, The 2019-new coronavirus epidemic: Evidence for virus evolution, J. Med. Virol., 92 (2020) 455-459.
12. F. Sievers, A. Wilm, D. Dineen, T.J. Gibson, K. Karplus, W. Li, et al, Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega, Mol. Syst. Biol., 7 (2011) 539.
13. Y. Zhang. I-TASSER server for protein 3D structure prediction, BMC Bioinformatics, 9 (2008) 40.
14. E.F. Pettersen, T.D. Goddard, C.C. Huang, G.S. Couch, D.M. Greenblatt, E.C. Meng et al. UCSF Chimera--a visualization system for exploratory research and analysis, J. Comput. Chem., 25 (2004) 1605-1612.
15. B. Shanmugaraj, K. Siriwattananon, K. Wangkanont, W. Phoolcharoen, Perspectives on monoclonal antibody therapy as potential therapeutic intervention for Coronavirus disease-19 (COVID-19), Asian Pac. J. Allergy Immunol., 38 (2020) 10-18.
16. J. Shang, G. Ye, K. Shi, Y. Wan, C. Luo, H. Aihara et al, Structural basis of receptor recognition by SARS-CoV-2, Nature, 581(2020) 221-224.
17. J.T. Ortega, M.L. Serrano, F.H. Pujol, H.R. Rangel, Role of changes in SARS-CoV-2 spike protein in the interaction with the human ACE2 receptor: an in silico analysis, EXCLI J., 19 (2020) 410-417.
18. C. Garcia-Iriepa, C. Hognon, A. Francés-Monerris, I. Iriepa, T. Miclot, G. Barone, et al, Thermodynamics of the interaction between SARS-CoV-2 spike protein and human ACE2 receptor effects of possible ligands, ChemRxiv., (2020) https://doi.org/10.26434/chemrxiv.12186624.v2.
19. G.A. Versteeg, P.S. van de Nes, P.J. Bredenbeek, W.J. Spaan, The coronavirus spike protein induces endoplasmic reticulum stress and upregulation of intracellular chemokine mRNA concentrations, J. Virol., 81 (2007) 10981-10990.